

Relative Contributions of Hudson River and Chesapeake Bay Striped Bass Stocks to the Atlantic Coastal Population

W. VAN WINKLE, K. D. KUMAR,¹ AND D. S. VAUGHAN²

*Environmental Sciences Division, Oak Ridge National Laboratory
Post Office Box 2008, Oak Ridge, Tennessee 37831-6038, USA*

Abstract.—Fourteen variables derived from 13 morphological characters were used in a stepwise discriminant analysis and a maximum-likelihood analysis to estimate the relative contributions of striped bass stocks from the Hudson River and Chesapeake Bay to the Atlantic coastal striped bass population. The analyses made use of the data collected by Texas Instruments in 1975, and were designed to focus on relative contributions by sex and year class to populations north of Chesapeake Bay and north of the Hudson River. The discriminant function method misclassified approximately 20% of the fish sampled on the spawning grounds. When applied to the data set for fish collected in the ocean, the two methods of analysis resulted in estimates of relative contribution of the Hudson stock to the Atlantic coastal population that varied considerably among year classes. In particular, the estimated relative contribution for the 1965 year class was between 40 and 50%, whereas the relative contributions for the 1966, 1968, and 1969 year classes were 10% or less.

One of the major issues in the Nuclear Regulatory Commission (NRC) licensing hearings for operation of Indian Point units 2 and 3 was the relative stock composition of the Atlantic coastal striped bass population. If Hudson River fish were major contributors to the coastal population, and if entrainment and impingement mortality of young-of-the-year striped bass at power plants along the Hudson were high, the Indian Point facility might contribute substantially to a reduction in the commercial and sport fisheries for striped bass along the Atlantic coast. In response to this concern, Consolidated Edison Company of New York funded a study by Texas Instruments (Grove et al. 1976; Berggren and Lieberman 1978).

Conventional wisdom, based on indirect evidence from commercial fisheries catch data and recoveries of tagged fish, has been that 90% or so of the legal-sized striped bass in the population migrating up and down the Atlantic coast from Maine to Cape Hatteras, North Carolina, comes from stocks spawning in the Chesapeake Bay and its tributaries. This wisdom was supported by the results of the Texas Instruments study, in which Berggren and Lieberman (1978) used observations of morphological characters and discriminant

analysis to quantitatively estimate the relative contribution of striped bass stocks from various estuaries to the striped bass fishery along the Atlantic coast. These authors estimated relative contributions of 91% from the Chesapeake Bay, 6% from the Hudson River, and 3% from the Roanoke River, North Carolina. These estimates were based on a sample of 2,471 ocean fish that included males and females of all ages from all 10 geographical strata and all 6 temporal strata defined in that study.

Our hypothesis was that the relative contributions are likely to vary substantially from year class to year class. We recognized the problems with limited sample sizes for certain year classes, but still believed that the range of year-class variation in relative contribution could be estimated from the original Texas Instruments data set. In 1977, we obtained on tape the complete data set for this Texas Instruments study. We repeated the Texas Instruments analysis and obtained identical results. Then we developed and applied alternative methods of analysis and focused on estimating relative contribution by sex and year class. We argue that the time is propitious to repeat this study.

Methods

Spawning-stock data.—The collection and processing of spawning-stock specimens was described by Berggren and Lieberman (1978). Briefly, mature striped bass were collected from the natal rivers (Figure 81) of major stocks along

¹Present address: SAS Institute Incorporated, Box 8000, Cary, North Carolina 27511, USA.

²Present address: National Marine Fisheries Service, Southeast Fisheries Center, Beaufort Laboratory, Beaufort, North Carolina 28516, USA.

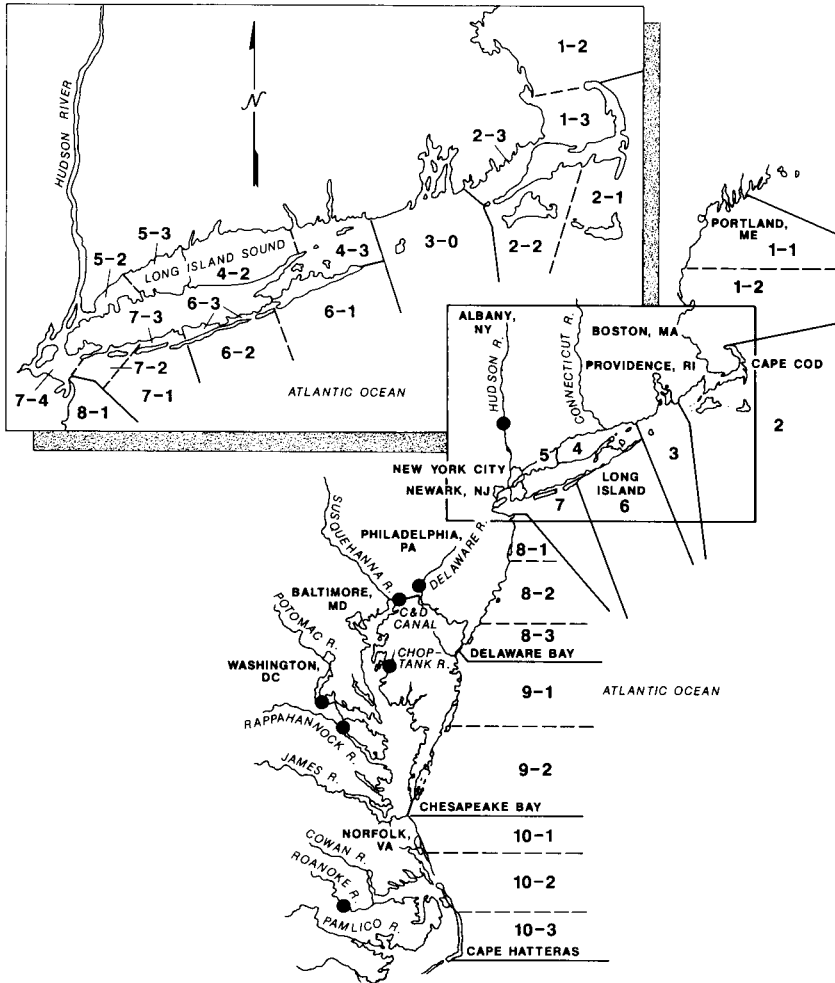


FIGURE 81.—Collection regions for the Atlantic coastal populations of striped bass, showing geographical stratification and substratification. Collection sites for spawning-stock specimens are indicated by dots on source rivers (Berggren and Lieberman 1978).

the Atlantic coast during the spawning season of 1975. Sampling was designed to obtain nearly equal numbers of male and female striped bass and a minimum of 10 individuals in each of several length categories in order to assure an adequate representation of the sexes and multiple year classes in the spawning-stock collections. The following 13 counts and measurements were made for each fish: number of lateral line scales, number of left pectoral rays, number of right pectoral rays, number of second dorsal rays, number of anal rays, number of upper-arm gill rakers, fork length, snout length, head length, internostril width, distance from focus to first annulus of

scales, distance from focus to second annulus of scales, and age as determined from scale annuli. For our analysis, we deleted the Roanoke fish from the spawning-stock data set. Our reasoning was as follows. Tag-recapture studies do not indicate appreciable migration of Roanoke fish north of the entrance to Chesapeake Bay or appreciable migration of Hudson fish south of the entrance to Chesapeake Bay. The controversy concerning the relative stock composition of the ocean population concerns primarily the area north of Chesapeake Bay. By deleting the Roanoke fish, we assume that we are simplifying the data set in a manner that will more accurately

allow us to estimate what we are primarily interested in, without the complicating and confounding effects of including a third stock with its own differences among ages and between sexes. Consistent with our deletion of Roanoke fish from the spawning-stock data set is our deletion of all fish from stratum 10 (south of the entrance to Chesapeake Bay) from the ocean data set (Figure 81).

Berggren and Lieberman (1978) reported that the relative contributions of the Roanoke stock to stratum 1 (Pemaquid Neck Light on the coast of Maine south to Race Point Light at the tip of Cape Cod and including all of Cape Cod Bay) were 11.5% (9 of 82 fish) during May–June and 4.6% (3 of 58 fish) during July–August. The relative contribution of Roanoke stock to stratum 2 (Race Point Light south along the outer coast of Cape Cod to the Massachusetts–Rhode Island border) was 24.0% (20 of 82 fish) in September–October 1975. These results may reflect extensive migrations by the Roanoke stock, but they are at odds with the extensive tag-recapture data (Clark and Baldrige 1984) currently available for the Roanoke stock. An alternative, and we feel more likely, interpretation is that these results are artifacts of the discriminant analysis procedure.

To minimize bias due to sex and year-class differences in the characters, we deleted all sex and year-class combinations of fish sampled on spawning grounds if there were fewer than two fish for either the Hudson or Chesapeake stock (Table 80). We repeated the analysis with a criterion of fewer than one fish for either spawning stock, and there were no pronounced differences in the results. The resulting data set includes 4–8-year-old and 10–11-year-old males and 6–7-year-old and 9–11-year-old females; 28 of 164 fish were deleted from the Hudson stock and 78 of 231 fish from the Chesapeake stock.

Ocean data.—Collection and processing of the ocean specimens were described by Berggren and Lieberman (1978). The same counts and measurements were made on the oceanic striped bass as on the spawning-stock striped bass. Geographic stratification consisted of dividing the Atlantic coast from Maine to Cape Hatteras into 10 collection regions, some with substratification (Figure 81); temporal stratification consisted of dividing the calendar year into six 2-month periods. As indicated above, we deleted from the ocean data set all fish caught in stratum 10 (Table 80). We also deleted from the ocean data set all fish that did not belong to one of the sex and year-class combinations retained in the

spawning-stock data set (Table 80). This left us with 798 ocean-caught striped bass for an analysis that we call case A.

Because there was only one 5-year-old female in the Hudson spawning-stock sample, this sex–age combination was not retained in the ocean data set, resulting in the deletion of 1,123 fish or 44% of the total number of fish in the ocean sample. The low abundance of 5-year-old females in the Hudson spawning stock is not surprising because Hudson female striped bass become sexually mature at age 6 and older, whereas Chesapeake females become sexually mature at age 4 and older (Hoff et al. 1988, this volume). However, because of the dominance of 5-year-old females in the ocean sample, a separate analysis was done that included this sex–age combination in the spawning-stock and ocean data sets. The dominance in the ocean sample of 5-year-old striped bass, both female and male (Table 80), reflects the 1970 dominant year class produced in the Chesapeake.

Three other analyses (cases B, C, and D) were performed with the discriminant function method but not with the maximum-likelihood method (see next section). The objective of these analyses was to test hypotheses for specific geographic strata. These three analyses involved deleting fish from strata 8 and 9 (case B), from strata 5 and 7–4 (case C), and from strata 5, 7–4, 8, and 9 (case D).

Statistical methods.—The primary goal of the statistical analysis was to estimate the relative contribution (p_1) of the Hudson River striped bass stock to the Atlantic coastal striped bass population. We estimated p_1 by two independent methods, the discriminant function method and the maximum-likelihood method.

Using equations estimated from the spawning-stock data set for striped bass known to be from either the Hudson or the Chesapeake, the discriminant function method attempts to classify each of the ocean fish of unknown origin as belonging to one group or the other. The methods of estimating the discriminant function and the “confusion matrix” and the derivation of the equation for calculating the relative contribution from the confusion matrix are provided in the Appendix. The estimates of contribution obtained from the discriminant analysis were corrected for bias due to misclassification by using percentages of specimens from each spawning stock that were misclassified into the other spawning stock. This is the same procedure used by Berggren and Lieberman (1978).

TABLE 80.—Sex, year class, and age composition of striped bass collected by Texas Instruments in 1975 in the Hudson River and Chesapeake Bay (spawning-stock data set) and in the ocean (ocean data set).

Year class	Age (years)	Number of legal-sized fish ^a		Number of ocean fish ^{a,b}
		Hudson	Chesapeake	
Males				
1973	2	0	(3)	(4)
1972	3	0	(26)	(16)
1971	4	11	20	87
1970	5	17	64	196
1969	6	13	5	25
1968	7	2	2	6
1967	8	5	3	7
1966	9	(8)	0	(13)
1965	10	13	2	3
1964	11	7	3	5
1963	12	(2)	0	(1)
1962	13	(2)	0	(1)
Total		68 (12)	99 (29)	329 (35)
Females				
1973	2	0	0	(4)
1972	3	0	0	(100)
1971	4	0	(4)	(234)
1970	5	(1)	(29)	(1,123)
1969	6	9	10	166
1968	7	7	6	48
1967	8	(1)	(11)	(78)
1966	9	18	24	151
1965	10	17	5	38
1964	11	17	9	66
1963	12	(11)	(1)	(76)
1962	13	(1)	(1)	(7)
1961	14	0	0	(6)
1960	15	(1)	0	(2)
1959	16	0	0	(4)
1958	17	(1)	(3)	(3)
1955	20	0	0	(1)
Total		68 (16)	54 (49)	469 (1,638)
Both sexes				
Total		136 (28) ^c	153 (78) ^c	798 (1,673)

^aNumbers in parentheses indicate fish in that sex and year class combination that were deleted from the spawning-stock data set and the ocean data set for our analysis. Criterion for deletion was fewer than two fish from either the Hudson River or the Chesapeake Bay. Legal-sized striped bass are fish greater than or equal to 406.5 mm fork length.

^bAll 51 ocean fish in stratum 10 (south of the entrance to Chesapeake Bay) were deleted from the ocean data set. Five fish in strata other than stratum 10 were deleted because sex was undetermined. One fish in stratum 9 was deleted because it was collected in Chesapeake Bay rather than the ocean. The 2,471 (798 + 1,673) fish in this table plus the 57 deleted fish just listed sum to 2,528 fish, which is the total sample size in Table 5 of Berggren and Lieberman (1978).

^cThese totals do not include four fish from the Hudson and one from the Chesapeake, which accounts for the discrepancy with the 168 Hudson fish and 232 Chesapeake fish reported by Berggren and Lieberman (1978). We did not include these five fish because they were not assigned an age by Texas Instruments due to conflicting age estimates based on scale annuli (J. T. Lieberman, Texas Instruments Incorporated, personal communication).

The maximum-likelihood method treats the task of estimating p_1 as a problem in estimating the parameters of two or more normal distributions from a mixture of these normal distributions; p_1 is estimated directly without classification of individual fish (Odell and Basu 1976). As a result, there is no equivalent of the bias problem encountered with the discriminant function method, and thus there is no confusion matrix or need to adjust

maximum-likelihood estimates of p_1 . Details are given in the Appendix.

We decided not to transform any of the basic data on the 13 morphological characters because we could not test statistically whether or not any transformations gave a better or poorer fit to a multivariate normal distribution; such a distribution is assumed for both the discriminant function and maximum-likelihood methods. This agrees

with procedures followed by Berggren and Lieberman (1978).

Under ideal circumstances, one would like to conduct the analysis for each sex and year-class combination separately so that one could obtain a clearer picture of the contribution pattern. However, because the sample sizes were not sufficiently large to allow such an analysis, we took the alternative route of attempting to "correct" the data for these effects.

Consider a specific sex (S) and year class (Y). For this sex and year-class combination (SY), we obtain the mean of the j th morphological character, $\bar{v}_j^{(SY)}$, over both spawning-stock samples. Then, for each fish in this sex and year-class combination in either spawning-stock sample, we define the new character

$$Y_{ijk}^{(SY)} = v_{ijk}^{(SY)} - \bar{v}_j^{(SY)},$$

$v_{ijk}^{(SY)}$ is the original value of the j th character for the k th fish from the i th spawning stock in sex and year-class combination SY . This mode of correction is based on a linear model for the effect of sex and year class on each morphological character. Because the averaging is done over both spawning stocks, it is essential that one must have data from both sources for a given sex and year class. As a consequence, several sex and year-class combinations were dropped from the analysis due to lack of data for both spawning stocks. Table 80 shows the sex and year-class combinations used in this study.

Alternative methods of accounting for sex and year-class effects were explored, involving regression of each morphological character on age or fork length for males and females separately. We preferred the above equation because it was the simplest method and involved the fewest assumptions.

The utility of the discriminant function is maximized when only the most "discriminating" characters are used in the function. We felt that the relationship between the 13 measured characters is complex and not fully understood or known, so we included the squares and cross products of these measures in the discriminant function, resulting in a total of 104 possible independent variables. This procedure is analogous to the second-degree polynomial approach used in empirical response-surface methods (Cochran and Cox 1957). The variables included in the function were determined by the stepwise discriminant function method (SAS 1982). We permitted a square of a measured character or a cross product

of two characters to be in the model even if the character itself did not appear. This allows the discriminant function to be more general than the one permitted by the usual quadratic discriminant function method used by Berggren and Lieberman (1978), which requires that the measured character itself be selected for the model before its squares or cross products are considered. The same character variables selected for the discriminant function method were used for the maximum-likelihood method.

Results and Discussion

Discriminant Function and Confusion Matrix

The stepwise linear discriminant function procedure resulted in selection of 14 (Table 81) of the 104 character variables by criteria described in the Appendix. Berggren and Lieberman (1978) found that five variables best discriminated among Hudson, Chesapeake, and Roanoke stock. In decreasing order of importance they were (1) the ratio of snout length to internostril width; (2) the ratio of the distance between the first and second scale annuli to the distance between the focus and the first annulus; (3) the sum of rays in the left and right pectoral, second dorsal, and anal fins; (4) the number of upper-arm gill rakers (including rudimentary rakers); and (5) the number of scales along the lateral line. Although there is not a direct correspondence between our 14 variables and these 5 variables selected by Berggren and Lieberman, it is apparent that snout length, internostril width, distance between focus and first annulus, distance between first and second annuli, and numbers of rays in the various fins are the most discriminating morphological characters in both analyses.

The confusion matrix for the spawning-stock data showed that of the 136 striped bass from the Hudson River, 28 (21%) were misclassified as Chesapeake fish (Table 82). Of the 153 striped bass from the Chesapeake, 30 (20%) were misclassified as Hudson fish. These misclassification percentages are higher than might be desired, but they are about the same as those of Berggren and Lieberman (1978), who misclassified 23% (39 of 168 fish) of the Hudson fish to the Chesapeake Bay or Roanoke River and 32% (74 of 232 fish) of the Chesapeake Fish to the Hudson or Roanoke.

Some of this misclassification undoubtedly arose because real differences between stocks were confounded with differences due to sex and

TABLE 81.—Variables selected in the discriminant function as determined by stepwise discriminant analysis of Hudson River and Chesapeake Bay striped bass.

Variable number	Description	F ^a
1	Snout length	55.0
2	Fork length	49.3
3	Number of rays on left pectoral fin	12.3
4	Distance from focus to first annulus of scale	10.2
5	Distance from first annulus to second annulus of scale	10.5
6	Product of number of soft rays on second dorsal fin and number of scales along lateral line	5.88
7	Product of number of rays on left pectoral fin and number of scales along lateral line	4.98
8	InternostriI width	4.86
9	Product of number of rays on right pectoral fin and number of upper-arm gill rakers, including rudimentary rakers	3.61
10	Product of internostriI width and number of upper-arm gill rakers, including rudimentary rakers	3.53
11	Product of number of soft rays on anal fin and head length	3.78
12	Square of number of soft rays on anal fin	5.12
13	Square of number of soft rays on second dorsal fin	3.12
14	Product of number of soft rays on anal fin and internostriI width	3.87

^aF = value of the F-statistic to remove a character from the discriminant function. The larger the F value, the more important that character is as a discriminator between the two stocks.

age. For example, when we repeated the analysis using only 5-year-old males from the Hudson and Chesapeake spawning stocks, only 1 of 17 (6%) Hudson fish was misclassified to the Chesapeake and only 2 of 64 (3%) Chesapeake fish were misclassified to the Hudson. The reason for selecting 5-year-old males for this example is that sample size was largest for this sex-age combination (Table 80).

Estimates of Relative Contribution

Errors in estimating the relative contributions of the Hudson stock to the coastal population were judged by the absolute value of the difference between the estimated and true relative contributions in the spawning-stock data by sex-age combination (Table 83). Neither of the two statistical methods we used resulted in a consistently smaller error than the other. Error varied with sample size as expected; the larger the sam-

ple, the smaller the error tended to be. Except for 7-, 10-, and 11-year-old males (sample sizes of 4, 15, and 10 fish, respectively), the error was less than 10%; for half of the 12 sex-age combinations, it was less than 5%. If this study is repeated, we recommend that length categories and sample sizes be selected to minimize the chance of an error greater than 10% for each sex-age combination in the spawning-stock data.

Estimates of relative contribution of the Hudson stock to the Atlantic coastal population of striped bass north of Chesapeake Bay ranged from 0 to 79%, depending on year class, sex, and method of estimation (Figure 82). The results indicate marked differences among year classes, and the two methods of estimating relative contribution gave give similar results.

Because 85% of the striped bass in the ocean sample were female (Table 80), it is appropriate to pay particular attention to the *p*₁ values in Figure 82 for females. The two estimates for 10-year-old Hudson females are between 40 and 50%, suggesting that the 1965 year class from the Hudson was relatively strong. The estimates for 6-, 7-, and 9-year-old females are all less than 10%, and these estimates are based on reasonably large sample sizes. The estimates for 7-year-old males are also low, consistent with the estimates for females. The separate analysis that included 5-year-old (1970 year class) female striped bass in the spawning-stock and ocean data sets resulted in an estimate of relative contribution of 1% from the

TABLE 82.—Confusion matrix for the striped bass spawning-stock data.

Actual	Classified ^a		Total fish
	Hudson (j = 1)	Chesapeake (j = 2)	
Hudson (i = 1)	108 fish φ11 = 0.79	28 fish φ12 = 0.21	136
Chesapeake (i = 2)	30 fish φ21 = 0.20	123 fish φ22 = 0.80	153

^a φij = number in row *i* and column *j* divided by the total for row *i*.

TABLE 83.—Absolute value of the difference between the estimated and true relative contributions (p_1) of Hudson River striped bass in the 1975 spawning-stock data set. Values for this error are given by sex and year-class combination for p_1 values calculated by both methods of analysis.

Year class	Age (years)	Sample size ^a	Discriminant function method			Maximum-likelihood method		
			p_1		Error ^b	p_1		Error ^b
			Estimate	True		Estimate	True	
Males								
1971	4	31	0.433	0.355	0.078	0.375	0.355	0.020
1970	5	81	0.217	0.210	0.007	0.202	0.210	0.008
1969	6	18	0.789	0.722	0.067	0.808	0.722	0.086
1968	7	4	0.513	0.500	0.013	0.326	0.500	0.174
1967	8	8	0.720	0.625	0.095	0.593	0.625	0.032
1965	10	15	0.679	0.867	0.188	0.872	0.867	0.005
1964	11	10	0.844	0.700	0.144	0.858	0.700	0.158
Females								
1969	6	19	0.470	0.474	0.004	0.410	0.474	0.064
1968	7	13	0.450	0.538	0.089	0.500	0.538	0.038
1966	9	42	0.395	0.429	0.034	0.460	0.429	0.031
1965	10	22	0.814	0.773	0.041	0.767	0.773	0.006
1964	11	26	0.641	0.654	0.013	0.661	0.654	0.007

^aCalculated from Table 1.

^bError = estimate - true.

Hudson. This is consistent with the strong 1970 year class from the Chesapeake, the largest on record (Clark and Baldrige 1984).

That the relative contribution from the Hudson varies is to be expected, because the historical record does not indicate a marked tendency for dominant or weak year classes to occur in the Hudson and Chesapeake simultaneously (Klauda et al. 1980). Our analysis indicates that the relative contribution varies from less than 5% to as high as 40 to 50%. The upper bound is less certain than the lower bound because sample sizes were small for most Hudson year classes that seemed to make a large relative contribution.

The estimates of relative contribution in Figure 82 are for the area north of Chesapeake Bay (case A). Because few tagged Hudson striped bass have been recovered south of Sandy Hook, New Jersey (McLaren et al. 1981), we predicted that by deleting the ocean fish from strata 8 and 9 (case B), the resulting relative contribution values (p_1) would be greater than those for case A. Next, we noted that the geographic area encompassed by case A does not include any of the Chesapeake Bay, although case A does include areas analogous to Chesapeake Bay at the mouth of the Hudson River: New York Harbor, Raritan Bay, and the western half of Long Island Sound. Therefore, we predicted that by deleting the fish from strata 5 and 7-4 (case C), the resulting p_1 values would be lower than those for case A. In testing both of

these predictions we included only those sex-age combinations for which sample size was greater than 10 fish and the estimate of relative contribution was greater than 1% for case A.

The analyses supported both predictions (Table 84). Every sex-age combination that showed a substantial decrease in sample size between case A and cases B or C gave p_1 values that changed in the direction expected. For the 1964 year class of female striped bass, as an example, deletion of 12 fish collected between Chesapeake Bay and Sandy Hook (strata 8 and 9) caused the relative contribution of the Hudson stock to increase from 18% (case A) to 23% (case B). Such results have limited value of themselves, but their agreement with our expectations is reassuring.

Our final analysis involved deletion of strata 5, 7-4, 8, and 9 (case D), leaving the area from eastern Long Island up through New England. The purpose of this analysis was to obtain estimates of p_1 for a region where relative contributions are currently a topic of debate. Our findings (Table 84) include examples in which p_1 is lower than that for case A (e.g., 4-year-old males) and higher than that for case A (e.g., 10- and 11-year-old females). We conclude, as we did for case A, that the relative contribution of the Hudson stock to this smaller geographic area varies considerably between sexes and among year classes, ranging from less than 10% to as great as 40 to 50%.

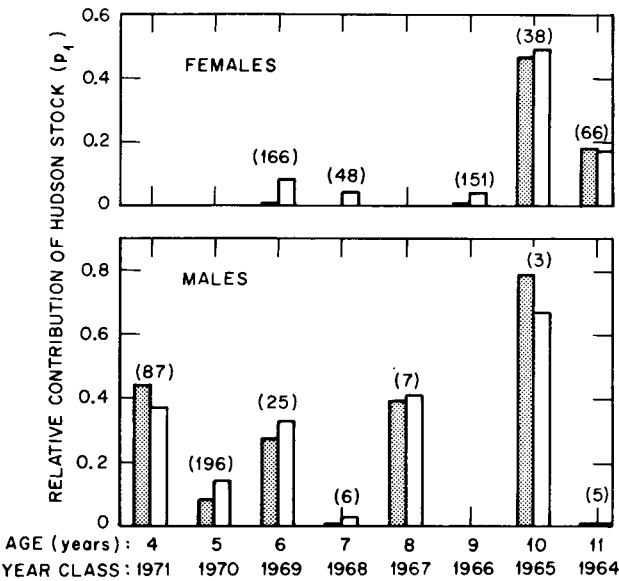


FIGURE 82.—Estimates from the 1975 ocean data of the relative contribution of the Hudson River striped bass to the coastal striped bass population north of Chesapeake Bay, for seven year classes of males and five year classes of females (sample sizes in parentheses), determined by the discriminant function method (stippled bars) and the maximum-likelihood method (clear bars).

Now that the 1970 dominant year class from the Chesapeake is no longer prominent in the ocean population, and no dominant year classes have appeared in the Chesapeake since 1970 (Clark and Baldrige 1984), the time seems propitious to repeat the study. This point is especially valid because the Hudson River population produced a good year class in 1978 and an exceptional year class in 1983 (Clark and Baldrige 1984). If the study is repeated, one implication of our further analysis of the 1975 Texas Instruments data is that the sampling design

should be modified to include more length categories, more fish per length category for each sex, or both, so that adequate numbers of striped bass are available to permit estimation of relative contribution for as many sex and year-class combinations as possible. Another implication is that possibly only those character variables found to be the most discriminating in both Berggren and Lieberman's (1978) and our discriminant analyses should be measured. This simplification would save money with little sacrifice in ability to estimate relative contributions.

TABLE 84.—Discriminant function estimates from the 1975 ocean data of the relative contributions (p_1) of the Hudson stock to the coastal striped bass population by sex and year-class combination for four geographic regions (cases A, B, C, and D).^a N denotes number of fish.

Year class	Age (years)	Case A		Case B		Case C		Case D	
		<i>N</i>	<i>p</i> ₁	<i>N</i>	<i>p</i> ₁	<i>N</i>	<i>p</i> ₁	<i>N</i>	<i>p</i> ₁
Males									
1971	4	87	0.441	82	0.452	57	0.205	52	0.222
1970	5	196	0.082	148	0.124	170	0.026	122	0.034
1969	6	25	0.274	21	0.379	20	0.181	16	0.296
Females									
1965	10	38	0.464	35	0.511	37	0.468	34	0.535
1964	11	66	0.179	54	0.232	66	0.185	54	0.232

^aRefer to Figure 81. Case A: region north of Chesapeake Bay; case B: strata 8 and 9 deleted; case C: strata 5 and 7-4 deleted; case D: strata 5, 7-4, 8, and 9 deleted, leaving eastern Long Island and New England.

Acknowledgments

We thank S. W. Christensen and J. A. Solomon for their critical reviews of this paper. The data analyzed were made available to us by Texas Instruments, Incorporated, through Consolidated Edison Company of New York, Incorporated, as part of the operating license proceedings for Consolidated Edison's Indian Point Nuclear Generating Plant, unit 3, before the U.S. Nuclear Regulatory Commission. This paper is based to a large extent on the report by Van Winkle and Kumar (1982). This research was supported by the U.S. Nuclear Regulatory Commission's Office of Nuclear Regulatory Research under interagency agreement DOE 40-550-75 with the U.S. Department of Energy under contract DE-AC05-84OR21400 with Martin Marietta Energy Systems, Incorporated. This is publication 2349 of the Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee.

References

- Anas, R. E., and S. Murai. 1969. Use of scale characters and a discriminant function for classifying sockeye salmon (*Oncorhynchus nerka*) by continent of origin. *International North Pacific Fisheries Commission Bulletin* 26:157-192.
- Berggren, T. J., and J. T. Lieberman. 1978. Relative contribution of Hudson, Chesapeake, and Roanoke striped bass, *Morone saxatilis*, stocks to the Atlantic coast fishery. *U.S. National Marine Fisheries Service Fishery Bulletin* 76:335-345.
- Clark, W., and M. Baldrige. 1984. Emergency striped bass research study, report for 1982-1983. Report to Congress by the Secretaries of Interior and Commerce, Washington, D.C.
- Cochran, W. G., and G. M. Cox. 1957. *Experimental designs*. Wiley, New York.
- Fournier, D. A., T. D. Beacham, B. E. Riddell, and C. A. Busack. 1984. Estimating stock composition in mixed stock fisheries using morphometric, meristic, and electrophoretic characteristics. *Canadian Journal of Fisheries and Aquatic Sciences* 41:400-408.
- Fukuhara, F. M., S. Murai, J. J. LaLanne, and A. Sribhibhadh. 1962. Continental origin of red salmon as determined from morphological characters. *International North Pacific Fisheries Commission Bulletin* 8:15-109.
- Gnanadesikan, R. 1977. *Methods for statistical data analysis of multivariate observation*. Wiley, New York.
- Grove, T. L., T. J. Berggren, and D. A. Powers. 1976. The use of innate tags to segregate spawning stocks of striped bass, *Morone saxatilis*. Pages 166-176 in M. Wiley, editor. *Estuarine processes*, volume 1. Academic Press, New York.
- Hoff, T. B., J. B. McLaren, and J. C. Cooper. 1988. Stock characteristics of Hudson River striped bass. *American Fisheries Society Monograph* 4:59-68.
- Kendall, M. G., and A. Stuart. 1973. *The advanced theory of statistics*, volume 2. Hafner, New York.
- Klauda, R. J., W. P. Dey, T. B. Hoff, J. B. McLaren, and Q. E. Ross. 1980. Biology of Hudson River juvenile striped bass. *Marine Recreational Fisheries* 5:101-123.
- Kshirsagar, A. M. 1978. *Multivariate analysis*. Dekker, New York.
- McLaren, J. B., J. C. Cooper, T. B. Hoff, and V. Lander. 1981. Movements of Hudson River striped bass. *Transactions of the American Fisheries Society* 110:158-167.
- Odell, P. L., and J. P. Basu. 1976. Concerning several methods for estimating crop acreages using remote sensing data. *Communications in Statistics. Part A: Theory and Methods* 5:1091-1114.
- Peters, C., and W. A. Coberly. 1976. The numerical evaluation of the maximum-likelihood estimate of mixture proportions. *Communications in Statistics. Part A: Theory and Methods* 5:1127-1135.
- Rao, C. R. 1952. *Advanced statistical methods in biometric research*. Wiley, New York.
- SAS (Statistical Analysis System). 1982. *SAS user's guide: basics*. SAS Institute, Cary, North Carolina.
- Tubbs, J. D., and W. A. Coberly. 1976. An empirical sensitivity study of mixture proportion estimators. *Communications in Statistics. Part A: Theory and Methods* 5:1115-1125.
- Van Winkle, W., and K. D. Kumar. 1982. Relative stock composition of the Atlantic coast striped bass population: further analysis. U.S. Nuclear Regulatory Commission, NUREG/CR-2563, Washington, D.C.

Appendix: Statistical Methods

Discriminant Function Method

Given a sample of N striped bass collected in the ocean, the discriminant function will classify \hat{N}_1 (the “^” denotes an estimate as opposed to the “true” value) as belonging to the Hudson River stock. The relative contribution of the Hudson stock (p_1) is given by

$$p_1 = \alpha \left(\frac{\hat{N}_1}{N} \right) + \beta; \quad (\text{A-1})$$

α and β are constants that can be estimated from the spawning-stock data set. In this section we discuss (1) the method of estimating the discriminant function, (2) the method for estimating the “confusion matrix,” and (3) the derivation of equation (A-1) from the confusion matrix.

Estimation of the linear discriminant function.—Let

$$y_1 = (y_{11}, y_{12}, \dots, y_{1K}) \text{ and } y_2 = (y_{21}, y_{22}, \dots, y_{2K})$$

be the vectors of K character variables for each fish (both sexes and all ages) sampled from the Hudson River and Chesapeake Bay, respectively. Let n_1 and n_2 be the numbers of fish sampled in each of the spawning stocks. We further define the following terms:

\bar{y}_i = mean sample vector of character variables for spawning stock i ($i = 1, 2$);

S_i = variance-covariance sample matrix for spawning stock i ($i = 1, 2$);

$n = n_1 + n_2$ = total sample size from the spawning stocks;

$\bar{y} = n_1\bar{y}_1 + n_2\bar{y}_2/n$ = overall mean sample vector of character variables;

$W = 1/n - 2[(n_1 - 1)S_1 + (n_2 - 1)S_2]$ = within-group, variance-covariance sample matrix;

$B = n_1(\bar{y}_1 - \bar{y})(\bar{y}_1 - \bar{y})' + n_2(\bar{y}_2 - \bar{y})(\bar{y}_2 - \bar{y})'$ = between-group, variance-covariance sample matrix, in which the prime denotes the transpose of the vectors of differences $(\bar{y}_i - \bar{y})$.

It is assumed that the y_i are samples from multivariate normal distributions and that the S_i are estimates of a common variance-covariance matrix. If the vector $z = a'y$ denotes a linear combination of the original character variables, a one-way analysis of variance for the derived variable z

will lead to the following F -ratio of the between-groups mean square to the within-group mean square:

$$F = \frac{a'Ba}{a'Wa}. \quad (\text{A-2})$$

If we choose the elements of the coefficient vector a such that this F -ratio is maximized, we are selecting the linear combination of the original character variables that best “discriminates” between the two stocks. The value of a is given by the eigenvector corresponding to the largest eigenvalue of $W^{-1}B$ (see Gnanadesikan 1977). Once the coefficient vector a has been determined, we can then classify the i th fish in the ocean sample as belonging to the Hudson stock if

$$a'x_i > a'(\bar{y}_1 + \bar{y}_2)/2; \quad (\text{A-3})$$

x_i is the vector of K character variables for the i th fish in the ocean sample. Otherwise, the i th fish is classified as belonging to the Chesapeake stock. The reader should refer to Rao (1952) and Gnanadesikan (1977) for more details.

The confusion matrix.—Once the discriminant function has been estimated, we can evaluate the effectiveness of the discriminant function by estimating the “confusion matrix.” Let ϕ_{ij} represent the proportion of the i th spawning stock that was classified as belonging to the j th spawning stock. Hence, ϕ_{11} and ϕ_{22} represent the proportions correctly classified as Hudson and Chesapeake, respectively, whereas ϕ_{12} and ϕ_{21} represent the proportions misclassified (e.g., ϕ_{12} represents the proportion of the Hudson spawning stock misclassified as Chesapeake spawning stock). The confusion matrix can be estimated by the jackknife method (also called the U -method; Kshirsagar 1978). The confusion matrix is estimated as follows:

Step (1). Compute the discriminant function for the two spawning-stock data sets combined, except for the i th fish.

Step (2). Classify the i th fish by the discriminant function computed in step 1.

We repeat the two steps for all the fish in the combined spawning-stock data set. The reader is referred to Kshirsagar (1978) for a general discussion of different methods of estimating the confusion matrix. The overall effectiveness of the discriminant function is then given by the ratio of the

total number of fish misclassified to the total number of fish in the spawning stock data set. The lower this number, the better the discriminant function.

*Estimation of relative contribution (p_1).—*The number of ocean fish classified as Hudson, N_1 , may be expressed as

$$\hat{N}_1 = \text{Prob}(1,1)N_1 + \text{Prob}(2,1)N_2; \quad (\text{A-4})$$

N_1 and N_2 are the true number of Hudson and Chesapeake fish in the ocean sample, respectively; $\text{Prob}(1,1)$ is the probability a Hudson fish is classified as Hudson, and $\text{Prob}(2,1)$ is the probability a Chesapeake fish is classified as Hudson. Dividing both sides by $N = N_1 + N_2$ gives

$$\hat{p}_1 = \text{Prob}(1,1)p_1 + \text{Prob}(2,1)(1 - p_1); \quad (\text{A-5})$$

\hat{p}_1 is the proportion of the ocean sample classified as Hudson, and p_1 is the "true" proportion of the ocean sample from the Hudson. Assuming ϕ_{11} equal to $\text{Prob}(1,1)$ and ϕ_{21} equal to $\text{Prob}(2,1)$ and solving for p_1 , we get

$$p_1 = \frac{\hat{p}_1 - \phi_{21}}{\phi_{11} - \phi_{21}}. \quad (\text{A-6})$$

Then

$$p_1 = \alpha \hat{p}_1 + \beta \quad (\text{i.e., equation A-1}); \quad (\text{A-7})$$

$$\alpha = 1/(\phi_{11} - \phi_{21}), \quad (\text{A-8})$$

and

$$\beta = -\phi_{21}/(\phi_{11} - \phi_{21}). \quad (\text{A-9})$$

This method of adjusting the estimates of contribution obtained directly from the discriminant analysis for bias due to misclassification is akin to moment estimation (Kendall and Stuart 1973). It is the same procedure used by Berggren and Lieberman (1978: their adjusted estimates); also see Fukuhara et al. (1962) and Anas and Murai (1969), who used this procedure to adjust estimates of stock contribution for sockeye salmon. If $p_1 < \phi_1$ or $\phi_{11} < \phi_{21}$ (but not both), the estimate of p_1 will be negative, indicating that the method is unable to estimate the contribution because it is very small or because the sample size is small. Whenever a negative p_1 was obtained, we set it equal to zero, unlike the procedure used by Berggren and Lieberman (1978).

Maximum-Likelihood Method

The maximum-likelihood method treats the task of estimating p_1 as a "mixture-of-normals" prob-

lem, and p_1 is estimated directly without classification of individual fish (Odell and Basu 1976; Peters and Coberly 1976; Tubbs and Coberly 1976; Fournier et al. 1984). As a result, there is no equivalent of the bias problem encountered with the discriminant function method and, thus, there is no need to adjust maximum-likelihood estimates of p_1 as in equation (A-7).

Let $f_1(y_1)$ and $f_2(y_2)$ be the K -dimensional probability density functions of the character variables for the Hudson stock and Chesapeake stock, respectively. The density function of the character variables for the ocean stock then is a mixture of these two spawning-stock density functions. In our case, we define for the ocean sample (denoted by x) a binomial distribution leading to the density function

$$f(x) = \hat{p}_1 f_1(x) + (1 - \hat{p}_1) f_2(x); \quad (\text{A-10})$$

\hat{p}_1 is the estimated contribution of Hudson stock to the Atlantic Ocean population. We further assume that $f_i(x)$ is a K -dimensional, multinormal distribution with mean vector θ_i and variance-covariance matrix Σ_i . Furthermore, to remain consistent with the linear discriminant function method, let $\Sigma_1 = \Sigma_2 = \Sigma$.

Given the spawning-stock data, we can readily obtain the usual estimates of θ_1 , θ_2 , and Σ . Hence, $f(x)$ can be re-written as

$$\hat{f}(x) = \hat{p}_1 \hat{f}_1(x) + (1 - \hat{p}_1) \hat{f}_2(x). \quad (\text{A-11})$$

One can then obtain the maximum-likelihood estimate of p_1 by forming the likelihood function of $f(x)$, differentiating the function with respect to p_1 , and setting this derivative equal to zero. After some algebraic manipulations we obtain the equation

$$\hat{p}_1 = \frac{\hat{p}_1}{N} \sum_{i=1}^N \frac{\hat{f}_1(x_i)}{\hat{p}_1 \hat{f}_1(x_i) + (1 - \hat{p}_1) \hat{f}_2(x_i)}; \quad (\text{A-12})$$

N is again the number of fish in the ocean sample. Because \hat{p}_1 occurs on both sides of the equation, a fixed-point solution method is used to estimate p_1 . The above equation is re-written as

$$\hat{p}_1^{(r)} = \frac{\hat{p}_1^{(r-1)}}{N} \sum_{i=1}^N \frac{\hat{f}_1(x_i)}{\hat{p}_1^{(r-1)} \hat{f}_1(x_i) + [1 - \hat{p}_1^{(r-1)}] \hat{f}_2(x_i)}; \quad (\text{A-13})$$

$\hat{p}_1^{(r)}$ = estimate of p_1 at the r th iteration.

The estimation algorithm is

Step (0). Let $\hat{p}_1^{(0)} = 0.5$. This is the initial estimate of p_1 .

Step (r). Substitute the estimate of p_1 at the $(r-1)$ th iteration ($\hat{p}_1^{(r-1)}$ in equation A-13) to obtain $\hat{p}_1^{(r)}$. If $|\hat{p}_1^{(r)} - \hat{p}_1^{(r-1)}|$ is less than some small number ϵ (0.00001 in our case), the algorithm has converged and $\hat{p}_1^{(r)}$ is our best estimate of p_1 . Otherwise we repeat step r.

It can be readily shown that, given $\hat{f}_1(x)$ and $\hat{f}_2(x)$,

the algorithm will converge. When $\hat{p}_1^{(r)}$ is close to zero or one, the algorithm is stopped, and $\hat{p}_1^{(r)}$ is set equal to zero or one, respectively.

One of the potential advantages of this maximum-likelihood method is that it is theoretically possible to calculate a confidence interval about \hat{p}_1 , whereas this is not possible with the discriminant function method. The procedure for doing this, however, would not be straightforward and would require developing additional computer programs.